# People Data Model

**Dated: 1/20/2016**          **Version: 0.1**

## Introduction

When we turn to the University for information, we expect it to be timely, thorough and, above all, accurate. However, we also demand the information furnished be split in many ways - academics, affiliations, event participation, demographics, location, university assets, etc. - representing data collected at different times and by different people.

A good example of data collection problems appears during a grant submission, where information about all aspects of individuals involved in research is gathered. Collectively, the University and its affiliates have hundreds of databases, with data stored in numerous formats, with new ones being added every year. Frequently, portions of this data overlaps or, in some cases, is complementary. Furthermore, the classes of data are often related or nearly identical.

In order for the University to standardize data types and enable effective sharing, a comprehensive and consolidated approach is necessary and should be outlined as:

- Create/Define a data dictionary[1] to enable the Columbia Community to locate the data they need easily and accurately.
- Create/Define data domains along with granular data attributes to support business processes and corresponding applications.
- Provide the ability to find and merge information, where applicable.
- Provide the ability to incorporate new data domains, where needed.
- Define a canonical data model[2] depicting inter-data and/or intra-data domain relationships for data sets.
- Support the use of Enterprise Integration Patterns[3] for automated data sharing, including web services, APIs and batch jobs.

## What is People Data?

People Data is any data about people affiliated with Columbia, including relevant attributes, roles, relationships to one another, utilization of University services, external entities, and so on. These people can include students, faculty, staff, alumni, patients, clients, family members, donors, prospective donors, prospective students, employees of vendors, contractors, community members, etc.

---

[1] https://en.wikipedia.org/wiki/Data_dictionary
[2] https://en.wikipedia.org/wiki/Canonical_model
[3] http://www.enterpriseintegrationpatterns.com/patterns/messaging/

# Phase I: Foundational Model

The goal of a foundational model is to help develop an understanding of fundamental university entities, including their relationships and their affiliations, in order to generate a logical data model for sharing information across the University. To achieve this goal, a People Data Working Group was formulated with representation from various disciplines across the university. Members contributed by submitting user stories from their corresponding business areas, which were analysed to identify the data attributes, data domain, Authoritative System of Record and relationships with other domains.

## Participating Business Units

- Alumni Development
- Center for Teaching and Learning
- College of Dental Medicine
- Columbia College
- CUIT
- CUMC IT Security
- Data Science Institute
- Facilities and Operations
- Graduate School of Art and Sciences
- Human Resources
- School of Continuing Education
- Teachers College

## Deliverables

The People Data Working Group collaborated to deliver the following deliverables:

- Data domains along with attributes.
- Canonical data model[4] depicting relationship between domains.
- System of Record (SOR) for data attributes.
- Process to maintain the people data model.
- These deliverables can be accessed [here](#).

## Process

The deliverables produced by the working group will be presented to the Architecture Assessment Group (AAG) and Architecture Steering Committee for review and comments. Upon successful review the deliverables are published as "People Data Model Standards".

---

[4] https://en.wikipedia.org/wiki/Canonical_model